GRAPHICAL MODELS AND IMAGE PROCESSING
Vol. 58, No. 5, September, pp. 405–412, 1996
ARTICLE NO. 0034

Why R.G.B.? Or How to Design Color Displays for Martians

P. GOLLAND AND A. M. BRUCKSTEIN*

Center for Intelligent Systems, Computer Science Department, Technion, I.I.T., Haifa 32000, Israel

Received July 20, 1995; revised June 17, 1996; accepted July 10, 1996

This paper deals with the problem of finding the best basis for color reproduction under a given model for color perception. We propose a mathematical framework for color basis optimization. We solve the problem for two different measures of "goodness" of a color basis and several perception models, including the human visual system. We also propose additional goddness measures for the design of special purpose displays. © 1996 Academic Press, Inc.

1. INTRODUCTION

Color is the perceptual response of a visual system to an input light with a spectrum defined by $S(\lambda) \geq 0$ for wavelength values λ in the visible range. The perception of color is due to detectors of three types present in the visual system, differing in their spectral sensitivity functions. These detectors will be called D_r, D_g, and D_b and their sensitivity or tuning functions will be denoted by $\varphi_r(\lambda)$, $\varphi_g(\lambda)$ and $\varphi_b(\lambda)$ accordingly. If all detectors are exposed to the same input spectrum $S(\lambda)$, the color perceived by the system [2, 5, 6, 8, 9] is determined by three nonnegative numbers, denoted by (r, g, b), obtained from the formulae

$$r = \int_\Omega S(\lambda)\varphi_r(\lambda)\, d\lambda,$$
$$g = \int_\Omega S(\lambda)\varphi_g(\lambda)\, d\lambda, \tag{1}$$
$$b = \int_\Omega S(\lambda)\varphi_b(\lambda)\, d\lambda,$$

the integration being over the visible range Ω.

The problem of color reproduction [6, 9, 11], using a mixture of pure colors is the following: we would like to choose a discrete set of wavelengths, $\{\lambda_1, \ldots, \lambda_k\}$, so as to be able to produce the widest range of, and, in the best case, all physically realizable nonnegative triplets (r, g, b). Those triplets encode the entire range of input spectra.

If (r, g, b) is a legal color triplet elicited by some spectrum $S(\lambda)$, we would like to find a discrete spectrum

$$S_D(\lambda) = \sum_{i=1}^{k} \alpha_i \delta(\lambda - \lambda_i) \tag{2}$$

that yields the same (r, g, b), so we require

$$r = \sum_{i=1}^{k} \alpha_i \varphi_r(\lambda_i),$$
$$g = \sum_{i=1}^{k} \alpha_i \varphi_g(\lambda_i), \tag{3}$$
$$b = \sum_{i=1}^{k} \alpha_i \varphi_b(\lambda_i), \tag{4}$$

or

$$\begin{bmatrix} r \\ g \\ b \end{bmatrix} = \begin{pmatrix} \varphi_r(\lambda_1) & \cdots & \varphi_r(\lambda_k) \\ \varphi_g(\lambda_1) & \cdots & \varphi_g(\lambda_k) \\ \varphi_b(\lambda_1) & \cdots & \varphi_b(\lambda_k) \end{pmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix} \triangleq \Phi\, \boldsymbol{\alpha}. \tag{5}$$

If the Φ matrix is full rank we could always choose three wavelenghts in terms of which the system will have a solution of the form

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{pmatrix} \varphi_r(\lambda_1) & \varphi_r(\lambda_2) & \varphi_r(\lambda_3) \\ \varphi_g(\lambda_1) & \varphi_g(\lambda_2) & \varphi_g(\lambda_3) \\ \varphi_b(\lambda_1) & \varphi_b(\lambda_2) & \varphi_b(\lambda_3) \end{pmatrix}^{-1} \begin{bmatrix} r \\ g \\ b \end{bmatrix} \triangleq \Phi_3^{-1} \begin{bmatrix} r \\ g \\ b \end{bmatrix}. \tag{6}$$

However the situation is not as simple as it seems to be, due to the fact that we are not able to generate negative coefficients $\{\alpha_i\}$, since we can add but never subtract photons of some required wavelength. So, instead of trying to

* On sabbatical at Bell Laboratories, Murray Hill, NJ 07974.

1077-3169/96 $18.00
Copyright © 1996 by Academic Press, Inc.
All rights of reproduction in any form reserved.

determine the possible triplets (r, g, b) that cause coefficients $(\alpha_1, \alpha_2, \alpha_3)$ be nonnegative, we will take another approach, namely given the turning curves $\{\varphi_r(\lambda), \varphi_g(\lambda), \varphi_b(\lambda)\}$ we want to determine three wavelength values $\{\lambda_1, \lambda_2, \lambda_3\}$ so as to have the largest "space" of $[r, g, b]^T$ vectors spanned by $\Phi_3[\alpha_1, \alpha_2, \alpha_3]^T$ with nonnegative coefficients $(\alpha_1, \alpha_2, \alpha_3)$.

There has been a lot of research done on the best way to reproduce color images [1, 6] given a certain set of basic colors. In this paper, we address a different problem: that of choosing the basic colors that will allow us to reproduce the richest possible palette. Even though the model we consider is a simplification of the models that take various practical aspects of cathode ray tube (CRT) colorimetry into account, it leads to interesting questions and provides a mathematical framework that can be extended to incorporate practical constraints.

The paper is organized as follows. In the next section we formulate the problem and propose two different measures of "goodness" of a color basis. In section 3 we solve the problem for both measures and for several color perception models, including the human sensitivity curves. Then we address related questions such as what we gain by using more than three basic colors and how the method can be extended to deal with nonpure colors. It is followed by a discussion on other possible measures of goodness of a color basis and conclusions.

2. PROBLEM DEFINITION

As defined in previous section, the problem is: given three sensors or detectors with sensitivity functions $\{\varphi_r(\lambda), \varphi_g(\lambda), \varphi_b(\lambda)\}$, find three wavelength values $\{\lambda_1, \lambda_2, \lambda_3\}$ so that the number of possible colors ((r, g, b) triplets) will be maximized. Before we start to solve this problem we need to precisely define a cost function to be maximized. This function must express user requirements for the color display. The CIE standard primary colors were chosen using the following simple requirement: the space of all feasible colors is to be maximized. In the two following sections we define and solve the problem with this particular requirement in mind. Two different cost functions are proposed, their properties are discussed, and the problem is solved for both of them. Then some other functions are considered.

Let us assume that we have chosen three wavelength values $\{\lambda_1, \lambda_2, \lambda_3\}$. All the coefficient triplets we can get using three display guns in a cathode ray tube with these frequencies (wavelength values) and maximal intensities I_1, I_2, I_3, form a parallelepiped spanned by three basis vectors:

$$\mathbf{e}_1 = \begin{bmatrix} I_1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ I_2 \\ 0 \end{bmatrix}, \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ I_3 \end{bmatrix}. \quad (7)$$

If Φ_3 is a full rank matrix, three vectors $\{\Phi\mathbf{e}_1, \Phi\mathbf{e}_2, \Phi\mathbf{e}_3\}$ are linearly independent and define a new parallelepiped that contains all the possible $(r, g, b)^T$ vectors that can be reproduced using the display guns defined above. For each possible coefficient triplet $(\alpha_1, \alpha_2, \alpha_3)$ and color perceived by a visual system is given by

$$\begin{bmatrix} r \\ g \\ b \end{bmatrix} = \Phi_3 \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \Phi_3 \left(\frac{\alpha_1}{I_1}\mathbf{e}_1 + \frac{\alpha_2}{I_2}\mathbf{e}_2 + \frac{\alpha_3}{I_3}\mathbf{e}_3 \right), \quad (8)$$

$$\begin{bmatrix} r \\ g \\ b \end{bmatrix} = \frac{\alpha_1}{I_1}(\Phi_3\mathbf{e}_1) + \frac{\alpha_2}{I_2}(\Phi_3\mathbf{e}_2) + \frac{\alpha_3}{I_3}(\Phi_3\mathbf{e}_3), \quad (9)$$

$$\begin{bmatrix} r \\ g \\ b \end{bmatrix} = \frac{\alpha_1}{I_1}\mathbf{b}_1 + \frac{\alpha_2}{I_2}\mathbf{b}_2 + \frac{\alpha_3}{I_3}\mathbf{b}_3, \quad (10)$$

where

$$\mathbf{b}_1 = \Phi_3\mathbf{e}_1, \quad \mathbf{b}_2 = \Phi_3\mathbf{e}_2, \quad \mathbf{b}_3 = \Phi_3\mathbf{e}_3 \quad (11)$$

form a basis of the perceived "color space." All the colors which can be reproduced by this gun system are linear combinations of vectors $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ with coefficients in interval [0, 1], so it will be natural to define our cost function as a volume of a parallelepiped spanned by the vectors $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$, which is proportional to the determinant of Φ_3:

$$f(\lambda_1, \lambda_2, \lambda_3) = I_1 I_2 I_3 \left| \det \begin{pmatrix} \varphi_r(\lambda_1) & \varphi_r(\lambda_2) & \varphi_r(\lambda_3) \\ \varphi_g(\lambda_1) & \varphi_g(\lambda_2) & \varphi_g(\lambda_3) \\ \varphi_b(\lambda_1) & \varphi_b(\lambda_2) & \varphi_b(\lambda_3) \end{pmatrix} \right|. \quad (12)$$

This is a possible cost function one should maximize in order to obtain an optimal wavelength set.

Another possibility is to consider a so-called "chromaticity diagram." If some input spectrum is perceived by a visual system as an (r, g, b) triplet, its normalized RGB parameters are defined by

$$\bar{r} = \frac{r}{r+g+b}, \bar{g} = \frac{r}{r+g+b}, \bar{b} = \frac{r}{r+g+b}. \quad (13)$$

These are not independent quantities—given two of them, the third one can be determined—so two independent parameters are enough to describe uniquely a point in the space of $[\bar{r}, \bar{g}, \bar{b}]^T$ vectors (which is actually a two-dimensional space). The chromaticity digram is a description of all realizable colors in, say, $(\bar{g}, \bar{r})$ coordinates.

Given the tuning curves, one can compute the pure color curve—a set of points corresponding to all pure colors on the chromaticity diagram. If three frequencies of the display guns are chosen, all the colors that can be reproduced using these guns are represented by points inside a triangle on a chromaticity diagram, whose vertices are points on the pure color curve corresponding to the wavelengths of the guns. Therefore, our cost function can be defined as the area of a triangle "spanned" by three points displaying the gun wavelength values on a chromaticity diagram.

Let us consider three points (P_1, P_2, P_3) on a chromaticity diagram corresponding to the gun wavelengths $(\lambda_1, \lambda_2, \lambda_3)$. The area of a triangle with vertices at these three points can be found from the formula

$$S = \|(\mathbf{P}_2 - \mathbf{P}_1) \times (\mathbf{P}_3 - \mathbf{P}_1)\|, \qquad (14)$$

so the cost of three given wavelength values $(\lambda_1, \lambda_2, \lambda_3)$ is defined by

$$g(\lambda_1, \lambda_2, \lambda_3) = \|((\bar{r}_2, \bar{g}_2) - (\bar{r}_1, \bar{g}_1)) \times ((\bar{r}_3, \bar{g}_3) - (\bar{r}_1, \bar{g}_1))\|, \qquad (15)$$

$$g(\lambda_1, \lambda_2, \lambda_3) = \left| \det \begin{pmatrix} \bar{r}_2 - \bar{r}_1 & \bar{r}_3 - \bar{r}_1 \\ \bar{g}_2 - \bar{g}_1 & \bar{g}_3 - \bar{g}_1 \end{pmatrix} \right|, \qquad (16)$$

where

$$\bar{r}(\lambda) = \frac{\varphi_r(\lambda)}{\varphi_r(\lambda) + \varphi_g(\lambda) + \varphi_b(\lambda)}, \bar{g}(\lambda) = \frac{\varphi_g(\lambda)}{\varphi_r(\lambda) + \varphi_g(\lambda) + \varphi_b(\lambda)}. \qquad (17)$$

It is important to clarify a major difference between the two cost functions defined above. While the first one takes into account both intensity and saturation with certain weights, the second one reflects only the richness of the display palette without consideration of the intensity range. So, their uses are different: if one can supply gun intensities high enough to get sufficient intensities for each color perceived by an observer, he would be interested in the second type of cost function. But, if it is also important to get good intensity ranges for the colors we are capable to reproduce (maybe even at the expense of not being able to reproduce some other colors at all), the first type of cost function is better suited. In the following discussion we will refer to both cost functions and consider optimal solutions for each one of them.

A further interesting question is how to choose more than three frequencies (wavelength) in order to extend the space of possible color triplets. A related issue is to determine the minimal number of pure colors necessary for covering the entire range of nonnegative 3-vectors encoding color.

3. SOLUTION

It is usually very difficult to solve the problem analytically in the general case (i.e., for any function set $\{\varphi_r(\lambda), \varphi_g(\lambda), \varphi_b(\lambda)\}$) and some numerical methods should be used in order to find the maximum of the function $f(\lambda_1, \lambda_2, \lambda_3)$. There are, however, simple cases when an analytical solution can be found. One of them is the "linear" case, discussed below. Two additional models with tuning curves defined analytically (polynomial and exponential, respectively) are considered in the Appendix. For both of them a cost function can be computed analytically, but its maximum is found by numerical computations because of its complicated analytical form. Then the tuning curves of the human visual system are analyzed and an optimal set of display gun wavelengths is determined. Its relationship with other standard color systems is discussed.

3.1. *Linear Model of Tuning Curves*

Assume that all the sensitivity functions have a linear form

$$\varphi_* = \begin{cases} J_* \dfrac{\lambda}{a_*} & 0 \le \lambda < a_* \\[2ex] J_* \dfrac{1-\lambda}{1-a_*} & a_* \le \lambda \le 1 \end{cases}, \qquad (18)$$

where $* \in \{r, g, b\}$, and $0 \le a_b \le a_g \le a_r \le 1$ (see Fig. 1a). In this case an exact analytical solution can be computed. The interval $[0, 1]$ should be divided into four subintervals: $[0, a_b)$, $[a_b, a_g)$, $[a_g, a_r)$, and $[a_r, 1]$, so that in each one of them the functions $\{\varphi_r(\lambda), \varphi_g(\lambda), \varphi_b(\lambda)\}$ are monotonic (increasing or decreasing) and have first order derivative at each inner point. Hence, if it is known to which subinterval each one of $\{\lambda_1, \lambda_2, \lambda_3\}$ belongs, an analytical form for f can be determined, then f is differentiated and the maximal value can be determined. An additional constraint is imposed by the assumption that $0 \le a_b \le a_g \le a_r \le 1$. So only those triples of the subintervals that satisfy this constraint are to be taken into account (e.g., a combination $\lambda_1 \in [a_b, a_g)$, $\lambda_2 \in [0, a_b)$, and $\lambda_3 \in [a_g, a_r)$ is impossible, because $\lambda_1 \le \lambda_2$). Here is an example for $\lambda_1 \in [0, a_b)$, $\lambda_2 \in [a_b, a_g)$, and $\lambda_3 \in [a_g, a_r)$:

$$f(\lambda_1, \lambda_2, \lambda_3) = I_1 I_2 I_3 \left| \det \begin{pmatrix} J_r \dfrac{\lambda_1}{a_r} & J_r \dfrac{\lambda_2}{a_r} & J_r \dfrac{\lambda_3}{a_r} \\[2ex] J_g \dfrac{\lambda_1}{a_g} & J_g \dfrac{\lambda_2}{a_g} & J_g \dfrac{1-\lambda_3}{1-a_g} \\[2ex] J_b \dfrac{\lambda_1}{a_b} & J_b \dfrac{1-\lambda_2}{1-a_b} & J_b \dfrac{1-\lambda_3}{1-a_b} \end{pmatrix} \right|;$$

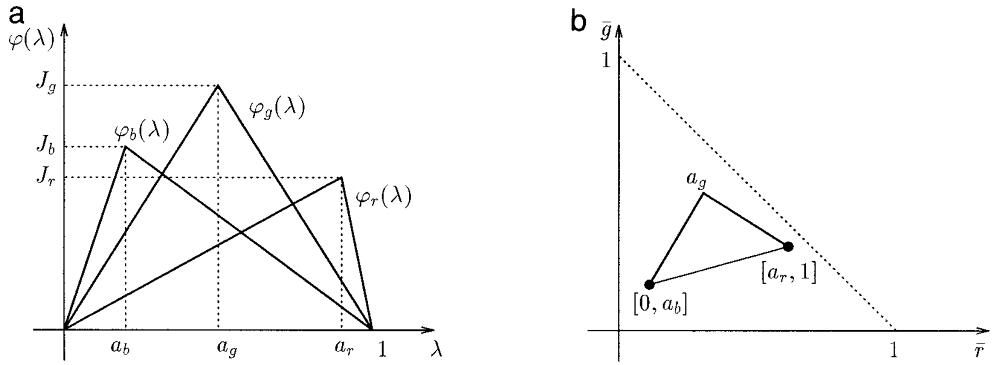

FIG. 1. A linear model of sensitivity curves: (a) sensitivity curves, (b) chromaticity diagram. On the chromaticity diagram the corresponding wavelengths are given.

$$f(\lambda_1, \lambda_2, \lambda_3) = \frac{I_1 I_2 I_3 J_r J_g J_b}{a_r a_b a_g (1 - a_b)(1 - a_g)} \lambda_1(\lambda_2 - a_b)(\lambda_3 - a_g).$$

It is obvious that f gets its maximal value when $\lambda_1 = a_b$, $\lambda_2 = a_g$, and $\lambda_3 = a_r$. All other combinations can be treated in a similar manner and yield the same result:

$$\lambda_1 = a_b, \lambda_2 = a_g, \lambda_3 = a_r. \tag{19}$$

This shows that for a linear model the points of a maximal sensitivity of the three different detector types should be taken as the wavelength values of the display guns.

Now let us build a chromaticity diagram for the linear model. For each of the four intervals mentioned above (r, g, b) values should be computed in order to find the corresponding points on the chromaticity diagram. After eliminating λ, we get (Fig. 1b)

$$\begin{cases} \bar{r} = \dfrac{p_r}{p_r + p_g + p_b}, \bar{g} = \dfrac{p_g}{p_r + p_g + p_b} & 0 \le \lambda \le a_b \\ \bar{g} = \dfrac{p_g}{p_r} \bar{r} & a_b \le \lambda \le a_g \\ \bar{g} = \dfrac{1}{1 + q_b q_g^{-1}} (1 - \bar{r}) & a_g < \lambda \le a_r \\ \bar{r} = \dfrac{q_r}{q_r + q_g + q_b}, \bar{g} = \dfrac{q_g}{q_r + q_g + q_b} & a_r \le \lambda \le 1 \end{cases} \tag{20}$$

where

$$p_* = \frac{J_*}{a_*}, q_* = \frac{J_*}{1 - a_*}, \tag{21}$$

for $* \in \{r, g, b\}$.

For the linear sensitivity model the pure color curve on a chromaticity diagram is a triangle with vertices at the points corresponding to maximum points of three tuning curves, so those three points represent an optimal solution for the second cost function. In order to get a maximal area of a triangle with vertices on the pure color curve, which is a triangle by itself, one should choose its vertices. The pure color curve vertices correspond to any triple $(\lambda_1, a_g, \lambda_2)$, where $0 < \lambda_1 \le a_b$ and $a_r \le \lambda_2 < 1$. Here one can clearly see a difference between two cost functions: any value of λ in the interval $[0, a_b]$ is equally good for the second cost function as a wavelength of the "blue" display gun, but the first one chooses a_b to be the best solution because it allows the widest range not only of color combinations, but also of color intensity. The same is true for the "red" display gun.

Although the linear model is a very rough approximation of human sensitivity curves, this result shows that it is advisable to choose the gun wavelength near the peaks of the receptor sensitivity curves, explaining the "engineering" choice of the RGB color basis. The computations and analysis are very similar in other analytical models. If analytical methods cannot be applied, numerical techniques are used to find the maximum point of a cost function. Two additional analytical examples are considered in the Appendix, and the next interesting model we explicitly analyze is the human vision model.

3.2. *Human Vision Model*

In previous examples the tuning curves were assumed to be available in an analytical form, but in most practical situations they are given in a discrete form (as a result of measurements). Tuning or sensitivity curves for the human vision system have been measured in laboratory conditions—the graphs are available in the literature [4, 7] and are shown in Fig. 2a. $\{\varphi_r(\lambda), \varphi_g(\lambda), \varphi_b(\lambda)\}$ have been determined for the wavelength range $\lambda \in$ [400 nm, 700 nm]

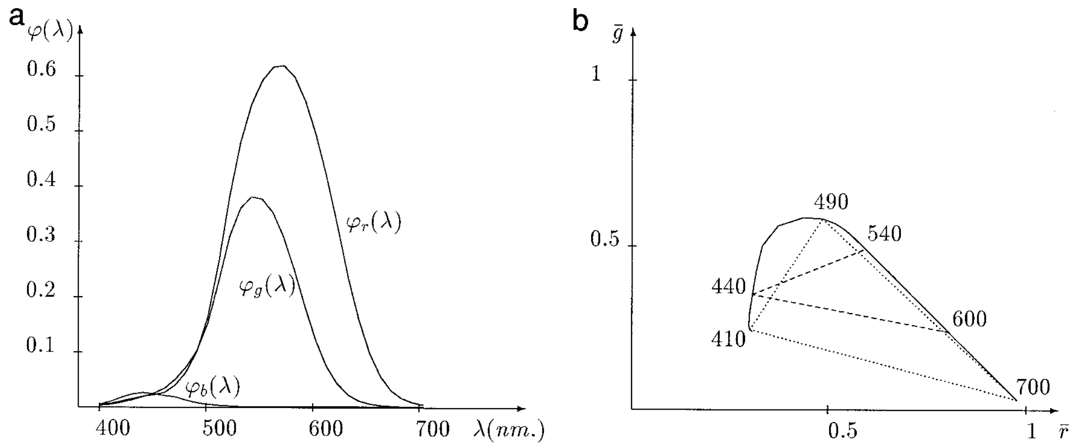

FIG. 2. Human vision model: (a) sensitivity curves, (b) chromaticity diagram.

with a step of 10 nm. For each node of a 3D-grid defined by 10 nm steps in the (λ_1, λ_2, λ_3) space the value of $f(\lambda_1, \lambda_2, \lambda_3)$ was computed and its maximum was found. The solution obtained is

$$\lambda_1 = 440 \text{ nm}, \lambda_2 = 540 \text{ nm}, \lambda_3 = 600 \text{ nm}, \qquad (22)$$

values that are, again, very close to sensitivity maximum points (440 nm, 540 nm, 570 nm) and correspond precisely to the three points of global maxima of the CIE standard primary color curves [3, 9]. Therefore in order to get the richest color set for humans by only three discrete frequencies, one should choose the global maxima of the CIE primaries. On the CIE chromaticity diagram all the colors that can be reproduced by such a system form a triangle with vertices located on the "horse shoe" curve, corresponding to the wavelength $\lambda_1 = 440$ nm (blue), $\lambda_2 = 540$ nm (green will a little of yellow), and $\lambda_3 = 600$ nm (orange).

In order to get the richest palette, i.e., maximize the second cost function $g(\lambda_1, \lambda_2, \lambda_3)$, the chromaticity diagram (Fig. 2b) was built and the maximal coverage area was found at

$$\lambda_1 = 410 \text{ nm}, \lambda_2 = 490 \text{ nm}, \lambda_3 = 700 \text{ nm}. \qquad (23)$$

These wavelength values allow us to get the widest range of color mixtures, and we can see that, again, these are the traditionally chosen colors (blue, green, red).

4. RELATED QUESTIONS

In view of the analysis done in the previous section, it is easy to answer several further questions about color reproduction.

4.1. *More Than Three Basic Colors*

For the first type of the cost function additional display guns cause the reproducible color space to be spanned by more than three vectors and it becomes more complex than a parallelepiped. Therefore addition of new pure colors extends the space of reproducible colors. The whole $[r, g, b]^T$ vector space, however, cannot be reproduced by any finite set of single wavelength guns, because, for example, the "pure" colors ($[1, 0, 0]^T$, $[0, 1, 0]^T$, and $[0, 0, 1]^T$) can never be reproduced by such a system. This is due to the fact that sensitivity functions are positive over the whole visible range, for the human visual system and for all other sensible models.

Considering the second type of the cost function, all reproducible colors $(\bar{r}, \bar{g}, \bar{b})$ are matched into inner points of the polygon with vertices on the pure color curve corresponding to gun wavelength. In the case of three basic colors, this polygon is a triangle, but in the general case it has a number of vertices equal to a number of guns used in the display system. Therefore the display palette can be enriched by the addition of new display guns, because the area of a convex polygon is larger then an area of a triangle generated by any three vertices of the polygon. A reasonable way to find a good solution for more than three display guns is to solve the problem for three guns and then to find an optimal location for each next gun wavelength by adding it to the solution found so far and optimizing this new problem. This is clearly a suboptimal, greedy approach.

We saw that for the linear problem a three gun system can cover the whole area inside the pure color curve. But for the human vision system and its 'smooth' model the pure color curve is smooth too, and it is impossible to cover an area with smooth borders by an polygon. So, no finite set of single wavelength display guns will be able to reproduce all possible colors inside the pure color curve.

4.2. *Nonpure Color Guns*

Another interesting question, which arises if practical issues of design are considered, is what happens if nonpure color guns are used. In real displays, phosphors with different spectral characteristics are used as light sources. They usually produce colors that are not pure, but rather a spectrum of wavelengths. The optimization process proposed above can also be applied to this case. In this case the set of the color primaries is a set of spectra $\{S_1(\lambda), \ldots, S_k(\lambda)\}$, which is a subset of all feasible phosphors, and Eq. (5) becomes

$$\begin{bmatrix} r \\ g \\ b \end{bmatrix} = \begin{pmatrix} \int_\Omega S_1(\lambda)\,\varphi_r(\lambda)\,d\lambda & \int_\Omega S_2(\lambda)\,\varphi_r(\lambda)\,d\lambda & \int_\Omega S_3(\lambda)\,\varphi_r(\lambda)\,d\lambda \\ \int_\Omega S_1(\lambda)\,\varphi_g(\lambda)\,d\lambda & \int_\Omega S_2(\lambda)\,\varphi_g(\lambda)\,d\lambda & \int_\Omega S_3(\lambda)\,\varphi_g(\lambda)\,d\lambda \\ \int_\Omega S_1(\lambda)\,\varphi_b(\lambda)\,d\lambda & \int_\Omega S_2(\lambda)\,\varphi_b(\lambda)\,d\lambda & \int_\Omega S_3(\lambda)\,\varphi_b(\lambda)\,d\lambda \end{pmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}. \quad (24)$$

A set of all possible spectra of display guns can be represented as a discrete set of vectors inside the "basic" parallelepiped defined by pure colors in the 3D space for the first cost function, and as a discrete set of points inside the pure color curve on the chromaticity diagram. The problem of determining an optimal color basis remains the same problem of maximizing the cost function, but now it is to be solved over a discrete set of possible solutions.

Trussel in [9] proposed to treat color spectra as multidimensional vectors (of about 30 dimensions). The vector components are the values of the spectrum sampled at small steps of λ. This approach was used in [10] for design of an optimal set of color filters. Using this approach in our case, it can be easily seen that Eq. (24) is a reduction of the multidimensional color space to its projection onto the 3D space spanned by the tuning curves $\{\varphi_r(\lambda), \varphi_g(\lambda), \varphi_b(\lambda)\}$. The case of nonpure color guns becomes almost identical to the case of pure color bases. The only difference between these two cases is that in the case of nonpure basic colors we have to actually perform an integration (inner product in the vector space of color spectra) in order to compute the various spectra's projection onto the space defined by the sensitivity functions.

5. OTHER COST FUNCTIONS

In the previous sections some very simple cost functions were proposed and analyzed. They reflect a requirement of uniform maximization of the feasible color space. But in some cases user demands may be different. In this section we discuss two examples of such demands and show how an appropriate cost function can be chosen in order to satisfy each of them.

Both functions proposed above consider their domain (3D space of vectors in the first case and 2D space of points in the second case) as a uniform one, but one could claim that some region, for example the region of colors around blue, is more important than another. If one defines a function of "importance" over the color space, the cost function could be defined not as the usual volume (area), but rather as a weighted integral over a parallelepiped (a triangle), the weight function measuring importance.

Another interesting example are special purpose displays, e.g., displays used for quality control on a production line. In this case only a limited group of materials will be displayed and their color properties are known at the time of design. One of the natural requirements for such displays is easy distinguishability between materials and the possibility to represent, as precise as possible, colors close to them. Therefore, the cost function in this case should maximize distances between the points representing given materials. The simplest solution is to find the minimal convex body (parallelepiped in the first case and triangle in the second) that includes all the points. In such a case the coefficients of different materials' representations will differ the most.

We can see that the proposed approach can be easily adjusted for various requirements, providing optimal choices of the basic colors for display guns.

6. CONCLUSIONS

Given three detector sensitivity curves, the problem of color reproduction using pure colors (i.e., three display guns with wavelength $\{\lambda_1, \lambda_2, \lambda_3\}$) can be solved by maximizing a cost function measuring the volume of a parallelepiped spanned by three vectors in the space of $[r, g, b]^T$ vectors, representing the color triplets produced by using the monochromatic display guns. Another natural cost function definition can be the area of a triangle with vertices corresponding to the three display gun wavelength values on the chromaticity diagram. These two functions differ in their treatment of the intensity and color mixture ranges. While the first one takes into consideration both color mixture and intensity ranges with some weights, the second one considers only color saturation without paying any attention to realizable gun intensities. The choice of a cost function, therefore, depends on the problem one is going to solve: if not only palette, but also a maximal intensity is important, the first cost function should be chosen. If the most important issue of some practical system is how rich its palette is, one should choose the second

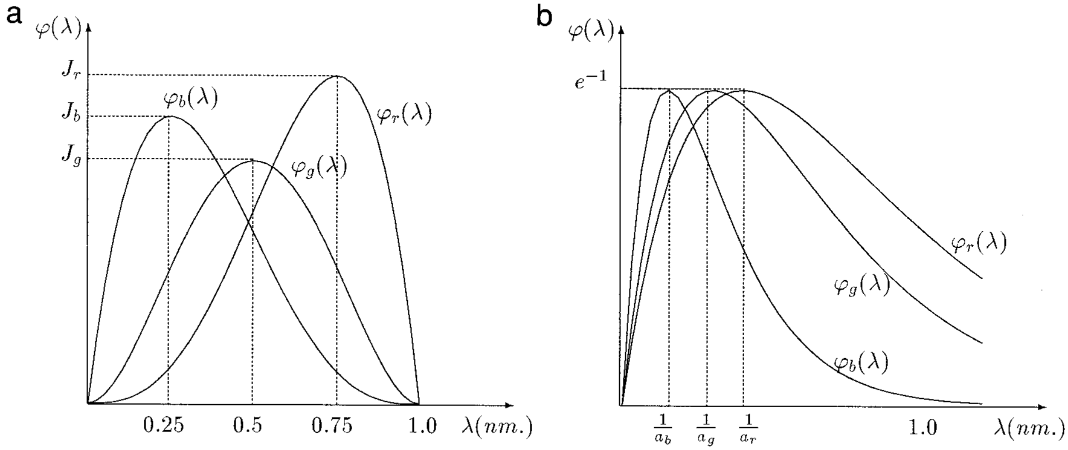

FIG. 3. (a) Polynomial and (b) exponential models of sensitivity curves.

approach. It is also possible to use a more complex cost function of the form

$$H(\lambda_1, \lambda_2, \lambda_3) = mf(\lambda_1, \lambda_2, \lambda_3) + (1 - m)g(\lambda_1, \lambda_2, \lambda_3), \quad (25)$$

where m defines the relative weights with which two previous types of cost functions are mixed together.

As we have seen, for some simple cases an analytical solution can be found, but in most practical situations (when tuning curves are given as a result of measurements, in a discrete form) numerical methods can readily be applied in order to maximize the cost function.

Addition of new guns with appropriate wavelength can extend a space of reproducible colors, but no finite set of guns can cover the whole range of possible colors.

The proposed method can be extended for nonpure color sources, as well as for more complicated requirements on properties of color display.

For the human vision system the optimal wavelength set is (not surprisingly) the three maximal points of the CIE standard "primary" color curves, i.e., the RGB basis. For Martians, the optimal set will probably comprise the three (or more) maximal points of the MCIE (Martian CIE) color curves.

APPENDIX

Polynomial and Exponential Models for Color Perception

Two models considered here are polynomial and exponential ones. They supply tuning curves in analytical form, but it is quite difficult to get a solution in such a form (due to the complexity of the derivatives). Therefore, both models were solved by numerical computations.

For each model the first cost function was maximized and an optimal wavelength triple was found.

6.1. *Polynomial Model*

Sensitivity curves are given by

$$\varphi_r(\lambda) = J_r\lambda^3(1 - \lambda), \quad (26)$$

$$\varphi_g(\lambda) = J_g\lambda^2(1 - \lambda)^2, \quad (27)$$

$$\varphi_b(\lambda) = J_b\lambda(1 - \lambda)^3, \quad (28)$$

where $\lambda \in [0, 1]$ (Fig. 3a). The cost function in this case

$$f(\lambda_1, \lambda_2, \lambda_3) = I_1I_2I_3\left|\det\begin{pmatrix} J_r\lambda_1^3(1-\lambda_1) & J_r\lambda_2^3(1-\lambda_2) & J_r\lambda_3^3(1-\lambda_3) \\ J_g\lambda_1^2(1-\lambda_1)^2 & J_g\lambda_2^2(1-\lambda_2)^2 & J_g\lambda_3^2(1-\lambda_3)^2 \\ J_b\lambda_1(1-\lambda_1)^3 & J_b\lambda_2(1-\lambda_2)^3 & J_b\lambda_3(1-\lambda_3) \end{pmatrix}\right|, \quad (29)$$

$$f(\lambda_1, \lambda_2, \lambda_3) = I_1I_2I_3J_rJ_gJ_b\lambda_1\lambda_2\lambda_3 (1-\lambda_1)(1-\lambda_2)(1-\lambda_3)(\lambda_2-\lambda_1)(\lambda_3-\lambda_2)(\lambda_3-\lambda_1). \quad (30)$$

By numerical methods one can find that $f(\lambda_1, \lambda_2, \lambda_3)$ achieves its maximal value at the point

$$\lambda_1 = 0.17, \lambda_2 = 0.5, \lambda_3 = 0.83. \quad (31)$$

This case can be also solved analytically.

6.2. *Exponential Model*

Let us consider tuning curves defined as follows

$$\varphi_r(\lambda) = a_r\lambda e^{-a_r\lambda}, \quad (32)$$

$$\varphi_g(\lambda) = a_g\lambda e^{-a_g\lambda}, \quad (33)$$

$$\varphi_b(\lambda) = a_b \lambda e^{-a_b \lambda}, \tag{34}$$

where $\lambda \in [0, 1]$. The cost function in this case is

$$f(\lambda_1, \lambda_2, \lambda_3) = I_1 I_2 I_3 \left| \det \begin{pmatrix} a_r\lambda_1 e^{-a_r\lambda_1} & a_r\lambda_2 e^{-a_r\lambda_2} & a_r\lambda_3 e^{-a_r\lambda_3} \\ a_g\lambda_1 e^{-a_g\lambda_1} & a_g\lambda_2 e^{-a_g\lambda_2} & a_g\lambda_3 e^{-a_g\lambda_3} \\ a_b\lambda_1 e^{-a_b\lambda_1} & a_b\lambda_2 e^{-a_b\lambda_2} & a_b\lambda_3 e^{-a_b\lambda_3} \end{pmatrix} \right|, \tag{35}$$

$$f(\lambda_1, \lambda_2, \lambda_3) = I_1 I_2 I_3 a_r a_g a_b \lambda_1 \lambda_2 \lambda_3 \left| \det \begin{pmatrix} e^{-a_r\lambda_1} & e^{-a_b\lambda_2} & e^{-a_r\lambda_3} \\ e^{-a_g\lambda_1} & e^{-a_g\lambda_2} & e^{-a_b\lambda_3} \\ e^{-a_b\lambda_1} & e^{-a_b\lambda_2} & e^{-a_b\lambda_3} \end{pmatrix} \right|. \tag{36}$$

Again, numerical methods are readily applicable in solving the problem. For example, for $a_r = 8$, $a_g = 10$, $a_b = 16$ (Fig. 3b), the three wavelength values should be

$$\lambda_1 = 0.04,\ \lambda_2 = 0.15,\ \lambda_3 = 0.36. \tag{37}$$

REFERENCES

1. R. S. Berns, R. J. Motta, and M. E. Gorzynski, CRT colorimetry, *Color Res. Appl.* **18**(5), 1993, 299–325.
2. G. Buchsbaum, A spatial processor model for object color perception, *J. Franklin Inst.* **310,** 1980, 1–26.
3. J. D. Foley, A. Van Dam, S. K. Feiner, and J. F. Hughes, *Computer Graphics: Principles and Practice,* 2nd ed., Addison–Wesley, Reading, MA, 1990.
4. P. Gouras, Color vision, in *Principles of Neural Science* (E. R. Kandel, Ed.), pp. 465–480, Prentice Hall International, London, 1991.
5. A. C. Hardy and F. L. Wurzburg, The theory of three-color reproduction, *J. Opt. Soc. Am.* **27,** 1937, 227–240.
6. B. K. P. Horn, Exact reproduction of colored images, *Comput. Vision Graphics Image Process.* **26,** 1984, 135–167.
7. E. H. Land, The retinex theory of color vision, *Sci. Am.* **237**(6), 1977, 108–128.
8. Q.-T. Luong, Color in computer vision, in *Handbook of Pattern Recognition and Computer Vision* (C. H. Chen, L. F. Pau, and P. S. P. Wang, Eds.), Chapt. 2, pp. 311–368, World Scientific, Singapore, 1993.
9. H. J. Trussel, DSP solutions run the gamut for color systems, *IEEE Signal Process. Magazine* Apr. 1993, 8–23.
10. P. L. Vora and H. J. Trussel, Measure of goodness of a set of color-scanning filters, *J. Opt. Soc. Am.* **10**(7), 1993, 1499–1508.
11. B. A. Wandell, The synthesis and anlysis of color images, *IEEE Trans. Pattern Anal. Mach. Intell.* **9**(1), 1987, 2–13.